AFRL-RI-RS-TR-2014-259

CAPTURING COGNITIVE FINGERPRINTS FOR ACTIVE AUTHENTICATION

IOWA STATE UNIVERSITY OF SCIENCE & TECHNOLOGY

OCTOBER 2014

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2014-259 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /
ANNA WEEKS
Work Unit Manager

/ S /
WARREN H. DEBANY, JR.
Technical Advisor, Information and
Exploitation and Operations Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
OCTOBER 2014	FINAL TECHNICAL REPORT	SEP 2013 – APR 2014

4. TITLE AND SUBTITLE

CAPTURING COGNITIVE FINGERPRINTS FOR ACTIVE AUTHENTICATION

5a. CONTRACT NUMBER FA8750-13-2-0088

5b. GRANT NUMBER N/A

5c. PROGRAM ELEMENT NUMBER 62722F

6. AUTHOR(S)

Jien Chang

5d. PROJECT NUMBER AAP2

5e. TASK NUMBER IO

5f. WORK UNIT NUMBER WA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Iowa State University of Science and Technology
1350 Beardshear Hall
Ames IA 50011-2025

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIGA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER

AFRL-RI-RS-TR-2014-259

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project intended to capture the cognitive fingerprints from individuals and use it as a biometric for continual authentication. This project proposed studying new biometric modalities for desktop (TA1a) and mobile devices (TA1b) in the first year and fusion methods for integration of modalities in the optional second year. For desktop computers, we developed Google Chrome plug-in to extract behavioral biometrics from mouse dynamics and web browsing behavior. The attribute of mouse biometrics is complementary to the one from keystroke (from phase 1 of this project) and offers great opportunity for integration in TA1a. For mobile devices, we developed an app containing 6 sub-apps with a variety of tasks to derive behavioral biometrics from gestures, virtual keyboards and sensors. The project ended before the IRB application was approved.

15. SUBJECT TERMS

Active Authentication, Cognitive Fingerprints, Biometric Modalities, Mouse Biometrics

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**ANNA WEEKS**
U	U	U	UU	11	**19b. TELEPHONE NUMBER** (*Include area code*) **N/A**

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

LIST OF FIGURES

Figures **Pages**

1.0 SUMMARY

Conventional authentication systems verify a user only during initial login. Active authentication performs verification continuously as long as the session remains active. This work focuses on using behavioral biometrics as “something a user is” for active authentication. This scheme performs continual verification in the background, requires no additional hardware devices, and is invisible to users.
This project intended to capture the cognitive fingerprints from individuals and use it as a biometric for continual authentication. This project intended to study new biometric modalities for desktop (TA1a) and mobile devices (TA1b) in the first year and fusion methods for integration of modalities in the optional second year. We developed data collection software applications and tools both for desktop and mobile devices, and we applied for IRB modifications in order to conduct the experiments necessary to obtain data that can be analyzed to extract the cognitive fingerprints. The project ended before the IRB application was approved.

2.0 INTRODUCTION

This project intended to study new biometric modalities for desktop (TA1a) and mobile devices (TA1b) in the first year and fusion methods for integration of modalities in the optional second year. We proposed to capture cognitive fingerprints from individuals and use them as biometrics for active authentication. For desktop computers, we proposed to extract behavioral biometrics from mouse dynamics. The attribute of mouse biometrics is complementary to the one from keystroke and offers great opportunity for integration in TA1a. For mobile devices, we planned to derive behavioral biometrics from gestures and virtual keyboards. Based on the experiences gained from our phase 1 project, we focus on the information induced by cognitive factors which have been ignored in past research.

We developed the software necessary to collect data both on desktop machines using browser extensions, and on mobile devices using an Android app. We intended to collect mouse and browsing behavior data from desktop machines. As for mobile devices, The app was programmed to collect touch screen virtual keyboard and gesture events as well as sensor data and web browsing behavior. This data was going to be analyzed but we needed to have IRB approval for conducting the experiments. The project was terminated before we were able to obtain IRB approval for our modifications.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

3.1 Desktop Data Collection

We worked on the design of experiments to collect mouse and web page browsing activities, implementation of web page interfaces for experiments, software design of web browser extension, pilot study to test the experiments, developing procedure and user interface (UI) for large scale experiment (including issues of recruiting, consenting, pilot testing and payment).

Tasks needed to accomplish this part included: (1) Design of experiments to collect mouse and

web page browsing activities, (2) Implementation of web page interfaces for experiments, (3) Software design for web browser extension, (4) Pilot study to test the experiment and (5) Development of procedure for large scale experiments.

The developed browser extension for data collection is shown in the following figures. Figure 1 (a) shows our plug-in was installed as one of the Chrome Extensions. We uploaded our plug-in to the Chrome Web Store so that participants could easily download it with the specific link which was given after they agreed to participate. After being successfully installed, our plug-in will lead the participants to an interface shown in Figure 1 (b). As we can see, the participants could still browse the web freely except there’s one extra black window designed by us, where the questions and instructions were shown to the participants. To minimize the disturbance to users’ regular browsing behavior, the task window could be minimized when users are browsing. Users will be asked to perform a series of tasks in approximately 30 minutes during each experiment session. This plug-in will only collect users’ mouse and web browsing behavior when they are during the sessions. After the experiments, the plug-in will be removed from the users’ computers automatically. The source code is attached to this report.

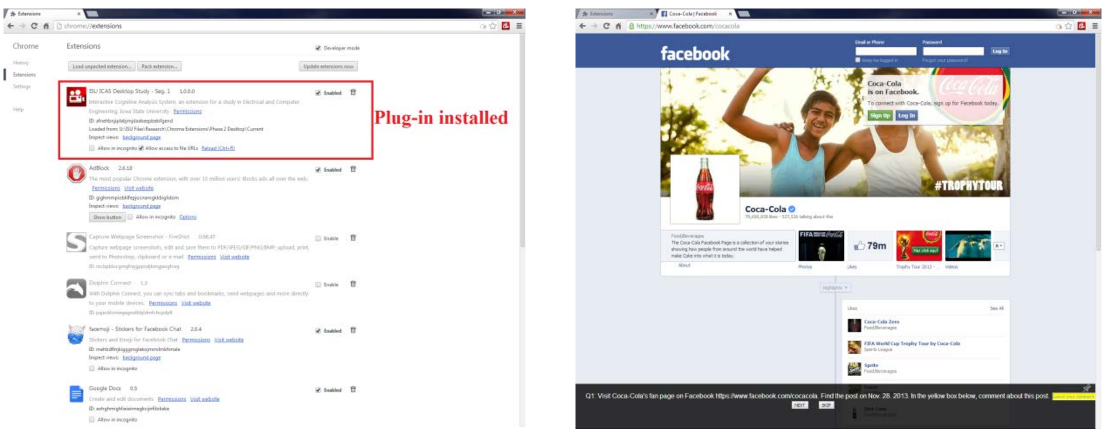

Figure 1. (a) The installed plug-in on Google Chrome. (b) The interface of our plug-in.

3.2 Mobile Data Collection

We worked on segments 1 & 2 of the mobile experiment which includes designing GUI of mobile app to collect gestures and virtual keyboard activities, software design of mobile apps, artwork design of mobile app, developing procedure and user interface (UI) for large scale experiment (including issues of recruiting, pilot testing and payment). Performing the pilot study and planning for further segments is halted because of the IRB issue.

Tasks needed to accomplish this part included: (1) Design of GUI of mobile app, (2) Artwork for the GUI, (3) Mobile GUI software structure, (4) Mobile software data (experiment details and questions), (5) Mobile app final integration, (6) Development of procedure for large scale experiment, (7) Mobile pilot study, and (8) Initial feature extraction from self-collected data for test purposes.

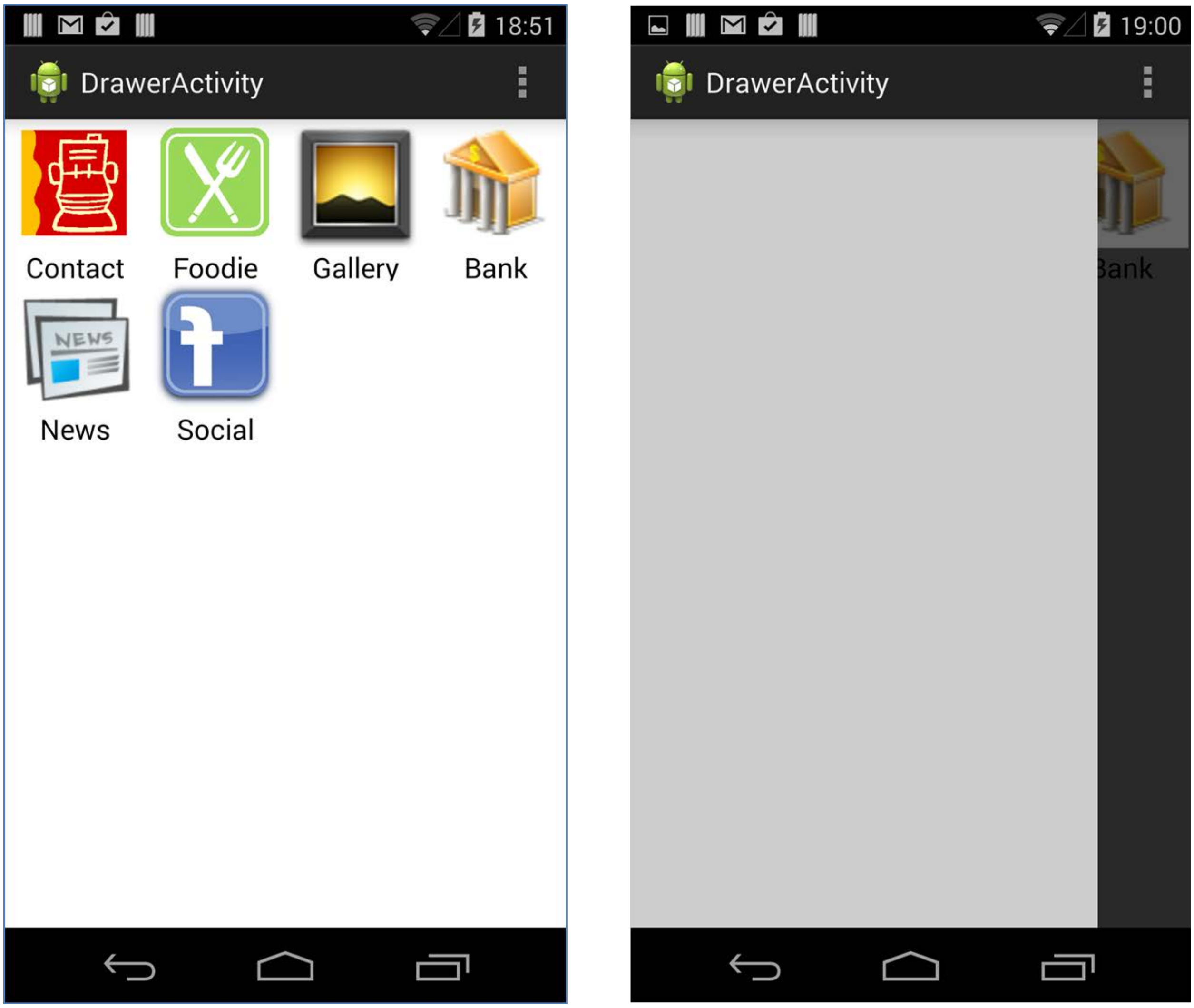

Figure 2. Two screenshots from the experiment app showing a list of sub-apps, and the experiment's instructions screen that shows up when the user swipes from left-to-right starting from the left edge.

There were six sub-apps that were specifically developed for this experiment (figure 2). They represented a wide variety of tasks that depict everyday usage of smartphones. These included sub-apps related to social networks, gallery, banking, news, restaurant reviews and phone contacts. Such diversity in tasks during the experiment would ensure that the participants would perform all kinds of gestures that are normally done on smartphones, and would also enable the research team to collect an adequate amount of data from virtual keystrokes. A screenshot of the developed app for data collection is shown in figure 3. The source code is attached to this report.

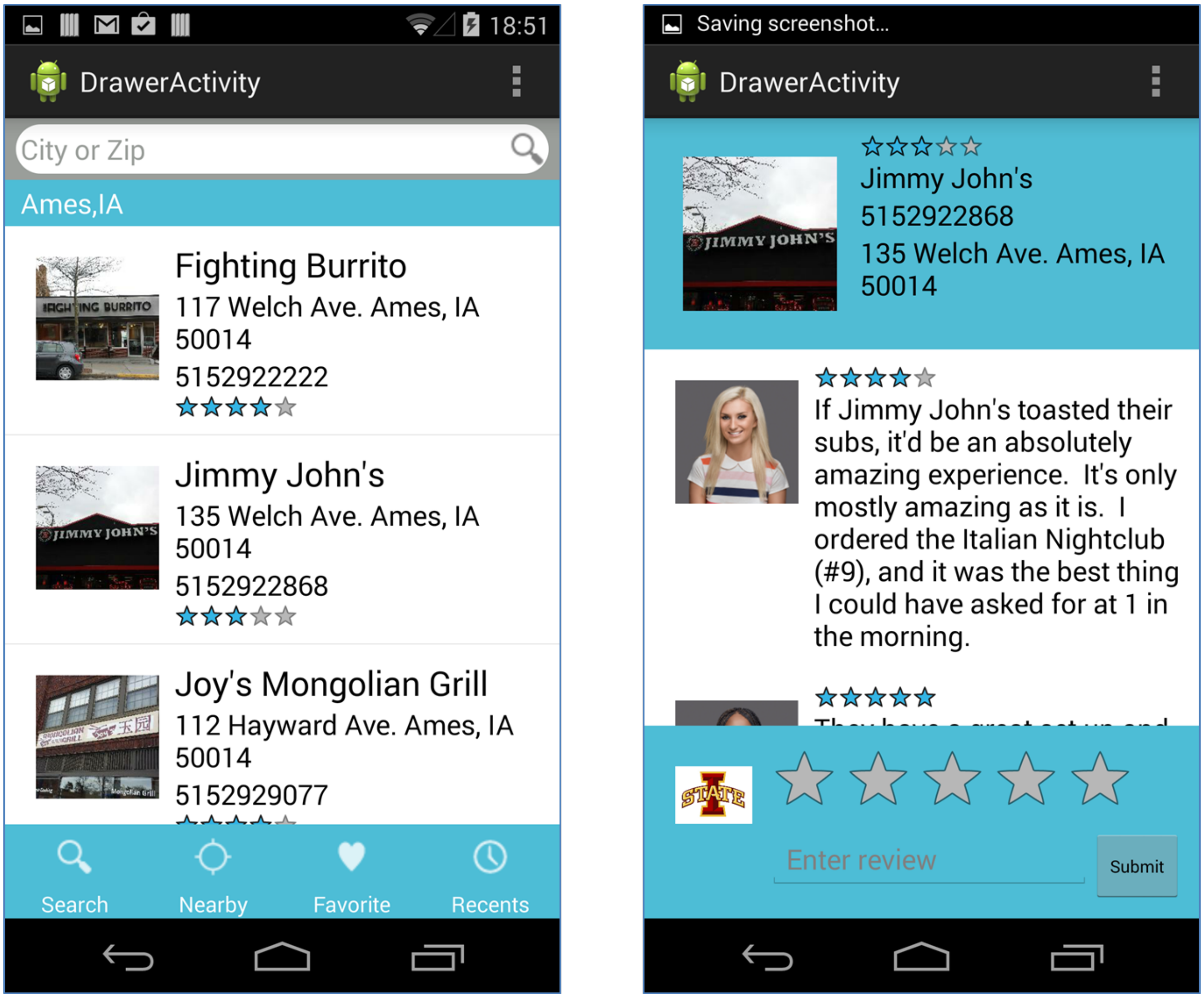

Figure 3. Two screenshots from the experiment app showing a list of restaurants, and a specific restaurant with customer reviews. All reviews were made by the research team.

3.3 Machine Learning / Fusion Methods

A novel truncated-RBF kernel was implemented to provide better cost-effectiveness tradeoff between computation cost and accuracy performance for learning algorithms. We continue to improve the machine learning techniques developed in phase 1. We also worked on designing more efficient learning algorithms to process the larger training data set, and to seek good tradeoff between classification performance and computation cost. We also surveyed previous work on various biometric machine learning methods, especially in website history, smart phone fingerprints, and mouse movement related applications.

We also looked into the decomposition of word into frequently used sub words to have more training samples, and to have useful n-graph information. Approximately 1000 prefix and suffix

sub-words were collected via Internet, and our existing system is being expanded to process them in addition to words.

Tasks needed to accomplish this part included: (1) Collecting frequently used prefixes/suffixes, (2) Code implementation that extends the existing system to take prefixes/suffixes into account, (3) Debugging and testing of the prefix/suffix functionality extension, (4) Testing the system with tri-graph sub-word features, (5) Surveying learning algorithms for various biometric features, (6) Supervised problem (Regression and classification problems), (7) Unsupervised problem (Clustering and data completion), and (8) Literature survey for algorithms dealing with incomplete data sets.

3.4 IRB Processing

We worked on some IRB modifications to conduct our large-scale experiments: (1) for long-term, it was first approved and then a hold was put on that until some issues related to the other set of IRB modifications are resolved, and (2) for the mobile and desktop large-scale experiments, we have submitted the minor modifications on Feb. 6th, and we expected it to be reviewed by the next IRB meeting on February 18th. However, the project terminated before obtaining an IRB approval (IRB modification document is attached)

4.0 RESULTS AND DISCUSSION

A paper titled “Cost-Effective Kernel Ridge Regression Implementation for Keystroke-Based Active Authentication System” was published in the proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [1].

The prefix/suffix ideas implemented in February does not improve the prediction accuracy. Nevertheless, with aid of the tri-graph sub-word features, the equal error rate (EER) of KRR algorithm (with TRBF2 kernel) is now reduced from 4.1% to 1.5%. Data collection didn’t happen as the project terminated before having approvals for our IRB application, and hence data analysis on new data wasn’t performed.

5.0 CONCLUSIONS

In summary, we improved the machine learning techniques that were developed in phase 1, and we improved the features we extracted to include tri-graph sub-word features that improved the ERR from 4.1% to 1.5%.

We have also developed the software needed to collect a wide variety of behavioral biometric features while providing the participants with user interfaces and tasks that best mimic their daily activities. This would help in extracting features that represent their normal behavior. We didn’t proceed to data collection though, because of IRB issues.

6.0 REFERENCES

1. Wu, P., Fang, C.; Chang, J.M.; Gilbert, S.B.; Kung, S.Y., "Cost-effective kernel ridge regression implementation for keystroke-based active authentication system," *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* , 4-9 May 2014, pp.6028-6032.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

EER	Equal Error Rate
GUI	Graphical User Interface
ICASSP	IEEE International Conference on Acoustics, Speech, and Signal Processing
IRB	Institutional Review Board
ISU	Iowa State University
KRR	kernel ridge regression
RBF	Radial Basis Function Kernel
TA1	Technical Area 1
TA2	Technical Area 2
TRBF2	Truncated Radial Basis Function Kernel Ver. 2
UI	User Interface